

Being Innovative Under ESSA's Innovative Assessment Demonstration Authority

An Innovator's Hope

Scott Marion

August 5, 2019

In my previous glass-half-empty [post](#), I outlined my considerable reservations with the Innovative Assessment Demonstration Authority (IADA) component of the Every Student Succeeds Act (ESSA).

Specifically, I expressed concerns with the comparability and scaling requirements of the Innovative Assessment Demonstration Authority because I see both of these requirements as serious threats to real innovation. Because I am really a glass-half-full kind of guy, I discuss here how we might address and overcome some of the technical, scaling, design, and policy obstacles to meaningful innovation in educational assessment.

Demonstrating the Technical Quality of Innovative Assessment Systems

Comparability remains a formidable technical challenge, but it is not the only technical quality requirement of the Innovative Assessment Demonstration Authority. Technical quality also includes requirements related to assessment quality, validity, reliability, and fairness. Other than the comparability concerns I outlined previously, I think we can meet the technical quality requirements, as long as the innovative assessments are not expected to be the same as more traditional state assessments. In other words, we can provide compelling validity evidence to support claims for the innovative assessments, but this evidence will look different than it does for statewide standardized tests. Colleagues and I outlined a comprehensive approach for documenting and [evaluating the technical quality](#) of innovative assessment systems, and we have many years of experience meeting these requirements for New Hampshire's [Performance Assessment of Competency Education](#), or PACE.

Challenge of Bringing Education Reform Programs to Scale is Not Limited to the Innovative Assessment Demonstration Authority

[Cynthia Coburn](#) (2003) has written about four interrelated dimensions that must be considered when thinking about bringing an education reform initiative to scale: depth, sustainability, spread, and ownership.

- **Depth** refers to changes that go beyond superficial structures and practices, such as simply replacing the state assessment with the innovative assessment without seeing fundamental changes in teaching and learning.
- **Sustainability** explicitly states that it is not enough for a reform to spread; it must be scaled with fidelity to allow the reform to be sustained.
- **Spread** emphasizes going beyond simply expanding to more districts to involve expanding in ways that pervade classrooms with the teaching and learning practices of the reform.
- "Finally, to be considered "at scale," **ownership** over the reform must shift so that it is no longer an "external" reform controlled by a reformer (i.e., the state and early adopters), but rather becomes an "internal" reform with authority for the reform held by districts, schools, and teachers who have the capacity to sustain, spread, and deepen reform principles themselves (Coburn, 2003, p.7)."

This conceptualization of scale does not make it any easier to meet the statewide scaling requirements of the IADA, but it can help states and districts plan for scaling. We offered additional suggestions for states and districts in this [brief](#), but going to scale remains the major challenge for innovative assessment reforms.

Balancing Innovation and Innovative Assessment Demonstration Authority Requirements through Thoughtful Design

One way states can resolve the inherent tension between innovation and the IADA requirements is by designing an innovative assessment system that considers the challenges of comparability and scaling from the beginning.

I am not suggesting backing off of innovation, but to perhaps think more in terms of a [balanced assessment system](#) approach that keeps the statewide summative assessment in place—albeit with a reduced presence. For example, one such approach might include a very short state assessment of about one class period for each content area. Such a short assessment would not support subscore reporting, but participating school districts would be required to administer and score 2-4 rich performance-based tasks tied to key components of the overall grade-level domain. The performance tasks can support deeper levels of learning and help fill out the alignment and subscore requirements.

Another balanced assessment system approach could rely on matrix-sampling for the statewide assessment (i.e., spreading out the full set of test questions among multiple test forms). Each student completes only part of the assessment, but the school receives information for the entire set of items. The school is the unit of analysis on the statewide assessment, but the ESSA requirements are clear that states must produce individual student total scores and subscores. Therefore, this matrix-sampling approach, combined with individual score information derived from local and common performance-based assessments, can fill out the assessment requirements—but in a way that supports improvements in teaching and learning.

These are just a couple of examples that spread the responsibility for technical quality to multiple assessments and do not just rely on a single form or type of assessment to meet all of the technical requirements of the Innovative Assessment Demonstration Authority. Colleagues came up with [15 possible designs](#) that states can consider as a starting point for design discussions. While we think these design templates could meet the federal requirements, any one of these designs must be tailored to a local context in an effort to address educational challenges in particular states and districts.

Working with Policymakers to Fill the Glass

I have tried to convey a more optimistic, glass-half-full tone in this post about how we might spur real innovation within the Innovative Assessment Demonstration Authority, but the fact remains that the innovative assessment glass contains half the amount of liquid possible. The Innovative Assessment Demonstration Authority was a good first attempt to introduce innovative assessment into federal education law, but like most first attempts at innovation, it fell a bit short. Thinking ahead to the next reauthorization of the Elementary and Secondary Education Act or even U.S. Department of Education (USED) waivers, here is my wishlist on how to fill the glass:

- The Innovative Assessment Demonstration Authority must come with some reasonable amount of funding—something on the order of \$1-2 million per state/year—so states can focus on the reform and not have to worry about raising money to support the reform.
- States should be permitted to put forth a limited number of innovations to provide a true laboratory.
- States should not be required to select a single innovation for scaling. This way, state and district leaders can tailor innovations to specific contexts and not have to worry about arbitrary timelines for scaling to a single statewide system.
- USED must engage with technical experts who understand what “comparable enough” means in order to meet the general assessment and accountability requirements of ESSA while still

allowing for innovation.

- Finally, states should be able to experiment with alternative accountability systems that hold schools appropriately accountable but do so in ways that capitalize on the different data the Innovative Assessment Demonstration Authority is generating. Subjecting pilot schools to the same accountability system as that of non-pilot schools is another serious limit to real innovation.

Whether you decide to drink from the Innovative Assessment Demonstration Authority glass and submit a Section 1204 application, or you choose to work outside of the Innovative Assessment Demonstration Authority to introduce innovative assessment in content areas or grade levels not subject to ESSA requirements, we need to commit ourselves to assessment innovations that support rather than hinder efforts to reform our teaching and learning systems.

I hope my suggestions help advance this goal.